


JOINT INVENTORS

"EXPRESS MAIL" mailing label No.

EK657816614US.

Date of Deposit: October 26, 2001

I hereby certify that this paper (or fee) is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 CFR §1.10 on the date indicated above and is addressed to: Commissioner for Patents, Washington, D.C. 20231


Laura Frasher

APPLICATION FOR UNITED STATES LETTERS PATENT

SPECIFICATION

TO ALL WHOM IT MAY CONCERN:

Be it known that we, Radoje T. Drmanac, a citizen of Yugoslavia, residing at 850 East Greenwich Place, City of Palo Alto and State of California, and Dennis G. Ballinger, a citizen of United States of America, residing at 38 Bishop Lane, City of Menlo Park and State of California and Ivan Labat, a citizen of Yugoslavia, residing at 1735 Steiner Street, in the City of San Francisco and State of California, and Birgit Stache-Crain, a citizen of Germany, residing at 345 South Mary Avenue, in the City of Sunnyvale and State of California, and Mark C. Dickson, a citizen of United States of America, residing at 1411 Gabilan Drive #B, in the City of Hollister and State of California, and Lee W. Jones, a citizen of United States of America, residing at 396 Ano Nuevo #412, in the City of Sunnyvale and State of California, and Aidong Xue, a citizen of People's Republic of China, residing at 1621 S. Mary Avenue, City of Sunnyvale and State of California,

and Y. Tom Tang, a Citizen of United States of America, residing at 4230 Ranwick Court, City of San Jose and State of California, and Chenghua Liu, a Citizen of People's Republic of China, residing at 1125 Ranchero Way #4, City of San Jose and State of California, and Vinod Asundi, a Citizen of United States of America, residing at 709 Foster City Blvd., City of Foster City and State of California have invented a new and useful STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS, of which the following is a specification.

4002949-100004

STEM CELL MAINTENANCE FACTOR MATERIALS AND METHODS

RELATED APPLICATIONS

This patent application claims priority to U.S. Serial No. 09/687,527 filed October 12, 2000, which is incorporated herein by reference in its entirety. This patent application is also a continuation-in-part of U.S. patent application Serial No. 09/378,667 filed August 20, 1999, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to a novel polynucleotide encoding a protein called hiwi-Hy, which is structurally related to a class of genes that are essential for stem cell self-renewal and maintenance, along with therapeutic, diagnostic and research utilities for these and related products.

BACKGROUND

The ability of stem cells to self-renew and to produce a large number of differentiated progeny is critical for the development and maintenance of a variety of tissues in organisms ranging from insects to mammals [Lin, H. *Annu. Rev. Genet.*, 31:455-491 (1997); Wylie, C., *Cell*, 96:165-174 (1999); Hogan, B., *Cell*, 96:225-233 (1999)]. This self-renewing ability is controlled by extrinsic and by cell-autonomous mechanisms [Lin, H. and Schagat, T. *Trends Genet.*, 13:33-39 (1997)]. Cell-autonomous mechanisms have been elucidated in a few stem cell models such as neuroblasts and germ line stem cells (GSCs) [Linn and Schagat, *supra*] whereas the role of extrinsic signaling (e.g. growth factors) in controlling stem cell division has been implicated in several systems [Morrison, S.J. *et al.*, *Cell*, 88:287-298 (1997)]. For example, the proliferation and differentiation of mammalian stem cells in hematopoietic, epidermal, and nervous systems depend on extrinsic signals that act on specific receptors on the stem cell surface [Morrison *et al. supra*]

Previous studies have shown that stem cell factor (SCF) and leukemia inhibitory factor (LIF) synergistically promote survival and in some cases the proliferation of mouse GSCs and primordial germ cells (PGCs) in culture [Dolci *et al. Nature* 352:809-811

(1991); Godin *et al.*, *Development*, 113:1451-1457 (1991); Matsui *et al.*, *Nature*, 353:750-752 (1991)]. However, under these conditions, PGCs have a finite proliferative capacity that correlates with their cessation of division *in vivo*. Recent studies have demonstrated that PGCs can be renewed and maintained in their undifferentiated totipotent state for some time in the presence of a cocktail of growth factors including β -fibroblast growth factor (β -FGF), SCF and leukemia inhibitory factor (LIF).

In addition to the extrinsic factors described above that are needed for the self-renewal and maintenance of PGCs, recent studies have identified three additional genes, *piwi*, *pumilio* (*pum*), and *fs(1)Yb*, which have been determined to be essential for maintenance and self-renewal [Lin, H. and Spradling, A.C., *Dev. Biol.*, 159:140-152 (1993); Lin, H. and Spradling, A.C., *Dev. Genetics*, 16:6-12 (1995); Lin, H. and Spradling, A.C., *Development*, 124:2463-2476 (1997)]. Among these genes, *piwi* is defined by recessive mutations that cause failure of PGCs maintenance in both males and females [Lin, H. and Spradling, A.C., *supra*]. In this study it was shown that the *piwi* gene is primarily involved in the promotion of self-renewing division of GSCs. *Piwi* is expressed primarily in somatic cells surrounding GSCs, and its function is required in those cells for GSC maintenance. In particular, in genetic mosaics where the surrounding somatic cells lack *piwi* gene function, there is a failure of self-renewing asymmetric cell divisions in adjacent normal GSCs, and existing germ cells degenerate [Cox *et al.*, *supra*]. Recent studies have demonstrated that *piwi* encodes a novel basic protein, and is well conserved among various species including humans [Cox D.N. *et al.*, *Genes and Develop.*, 12:3715-3727 (1998)]. Among this class of genes, the significantly higher homology between the *drosophila* gene *piwi* and its human homolog, *hiwi*, suggests that *hiwi* function is closer to *piwi*. Consistent with this observation, GSC division and gametogenesis in humans are much more similar to that in *Drosophila* whose gonads contain syncytial mitotic germ line cells that are capable of self-renewal only as a population. These studies demonstrate that there are a number of genes that control entry into the germ line and regulate the properties of germ cells during differentiation and development.

Primordial germ cells (PGCs) are thought to be derived from a small population of embryonic ectoderm cells set aside prior to gastrulation [Lawson and Pederson, *CIBA*

Foundation Symposium 165- Post Implantation Development in the Mouse, John Wiley and Sons (1992)] or even earlier [Soriano and Jaenish, *Cell*, 46:19-29 (1986); Wylie, *Cell*, 96:165-174 (1999)]. When they appear first in the embryo, germ cells have the potential to differentiate into gametes of either sex, depending upon the signals from the environment. These cells also carry the property of totipotency, i.e. the ability to differentiate into any cell type. These cells are the means by which species form and change in evolution. Germ line stem cells (GSCs) are derived from the primordial germ cells (PGCs) and provide a steady and continuous source of germ cells for the production of gametes with a remarkable totipotency to generate new individuals. Under certain conditions, PGCs give rise to teratomas and transplantable teratocarcinomas containing pluripotential embryonal carcinoma (EC) stem cells [Noguchi and Stevens, *J. Natl. Cancer Inst.*, 69:907-913 (1982)]. More recently, pluripotential embryonic stem (ES) cells have been isolated from the inner cell mass (ICM) from an embryo at the blastocyst stage [U.S. Patent No. 5,690,926]. Since PGCs, GSCs, and pluripotential ES cells can give rise to virtually any mature cell type, they are of great value for uses such as creating genetically manipulated animals and in tissue engineering.

Thus there exists a need for generating and maintaining stem cells in their undifferentiated, totipotential/pluripotential state. In addition, there exists a need for modulators of entry of GSCs into the germ line and modulators of the properties of GSCs during differentiation and development.

SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, in particular, a novel human stem cell maintenance factor and active variants thereof, isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically

engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

A nucleotide sequence encoding a portion of a stem cell maintenance factor designated hiwi-Hy is set forth in SEQ ID NO: 1, and the deduced amino acid sequence of the open reading frame (ORF) is set forth in SEQ ID NO: 2. The polypeptide set out in SEQ ID NO: 2 displays amino acid homology with the *Drosophila* piwi, as well as with human hiwi, *C. elegans* ciwi1, *C. elegans* ciwi2 and *Drosophila* aubergine. An alignment of a portion of hiwi-Hy (SEQ ID NO: 1) and other members of the stem cell maintenance factor gene family is shown in Figure 1. Hiwi-Hy shares a 54% amino acid identity with the human homolog, hiwi, 38% identity with stem cell maintenance factor homolog from *Caenorhabditis elegans*, and 34% identity with the stem cell maintenance factor homolog from *Drosophila melanogaster*. Additional family members can be identified using SEQ ID NO: 1 as a molecular probe. The polypeptide of SEQ ID NO:2 also displays amino acid homology with SEQ ID NO: 9, which is the same as the amino acid sequence of SEQ ID NO: 355 of U.S. Serial No. 09/687,527, filed October 12, 2000, incorporated herein by reference in its entirety, including but not limited to SEQ ID NO: 355. An alignment of SEQ ID NO:2 with a portion of amino acid sequence SEQ ID NO: 9 is shown in Figure 2. The alignment shows that the hiwi-Hy amino acid sequence (SEQ ID NO: 2) shares 98% amino acid identity with amino acid position 207 through amino acid position 392 of SEQ ID NO: 9.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. SEQ ID NO: 1 represents a portion of the coding region of the cDNA. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to SEQ ID NO: 1 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using SEQ ID NO: 1 or a portion thereof as a probe. Alternatively, SEQ ID NO: 1 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID

NO: 2 or the corresponding full length or mature protein sequence or a portion thereof.

The isolated polynucleotides of the invention further include, but are not limited to, a polynucleotide comprising the nucleotide sequence of SEQ ID NO: 1; a polynucleotide comprising the full length protein coding sequence corresponding to SEQ ID NO: 1 or a portion thereof; and a polynucleotide comprising the nucleotide sequence of the mature protein coding sequence corresponding to SEQ ID NO: 1 or a portion thereof. The

polynucleotides of the present invention also include, but are not limited to, polynucleotides that encode polypeptides with stem cell maintenance factor activity and that hybridize under stringent hybridization conditions to the complement of (a) the nucleotide sequence of SEQ ID NO: 1, or (b) a nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 2; a polynucleotide which is an allelic variant of any polynucleotide recited above; a polynucleotide which encodes a species homolog (ortholog) of any of the proteins recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptide corresponding to the amino acid sequence of SEQ ID NO: 2. The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising the amino acid sequence of SEQ ID NO: 2 or the corresponding full length or mature protein or a portion thereof. Polypeptides of the invention also include polypeptides with stem cell maintenance factor activity that are encoded by (a) polynucleotides comprising SEQ ID NO: 1; or (b) polynucleotides that hybridize to the complement of SEQ ID NO: 1 under stringent hybridization conditions. Biologically or immunologically active variants of the stem cell maintenance factor corresponding to SEQ ID NO: 2 and "substantial equivalents" thereof (e.g., with 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that retain stem cell maintenance factor activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also relates to methods for producing polypeptides of the invention comprising growing a culture of the cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the protein from the cells or the culture medium in which the cells are grown. Preferred
5 embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers for PCR, use for
10 chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect or quantify the presence of the particular cell or tissue mRNA in a sample using, e.g., in situ
15 hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath *et al.*, *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

20 The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the
25 invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject or cells from said subject a therapeutically effective amount of a composition comprising a protein of the present invention and a pharmaceutically acceptable carrier.

30 Where the polypeptide has stem cell maintenance factor activity, the polypeptides and polynucleotides of the invention can be utilized, for example, as part of methods for

the maintenance and/or renewal of stem cells, including embryonic stem cells (ES), germline stem cells (GSCs), primordial germ cells (PGCs), hematopoietic stem cells, including progenitor cells, epidermal stem cells, gut stem cells, and nervous system stem cells, or in methods for treating infertility. Where the polypeptide promotes cell differentiation, polypeptides and polynucleotides can be utilized, for example, to promote organ or tissue regeneration, to stimulate or inhibit gametogenesis, or to affect fertility.

Previous studies have shown that mutation in stem cell maintenance factor genes that abrogates the expression of these genes leads to a smaller number of gametes in the adult gonads that no longer contain GSCs (Lin, H and Spradling, A-C, *Supra*). This failure of germ line maintenance resulting from a mutation in the stem cell maintenance factor genes have been attributed to: (1) the differentiation of GSCs without self-renewing divisions; (2) a defect in the asymmetry of GSC division producing aberrant germ cells that eventually degenerate; and/or (3) a secondary defect influenced by abnormal ovary differentiation [Cox *et al.*, *supra*].

The methods of the present invention further relate to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders associated with the hiwi-Hy gene as described herein (e.g., fertility disorders, hematopoietic stem cell disorders or other stem cell disorders) and for the identification of subjects exhibiting a predisposition to such conditions. The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited herein.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention.

The methods of the invention also include methods for the treatment of disorders as recited above which may involve the administration of such compounds to individuals exhibiting symptoms or tendencies related to disorders as recited herein. In addition, the invention encompasses methods for treating diseases or disorders as recited herein

5 comprising the step of administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either on the level of target gene expression or target protein activity.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A-H shows an alignment of a portion of hiwi-Hy amino acid sequence (also called CG389 and set forth in SEQ ID NO: 2) with: human hiwi (SEQ ID NO: 3),

10 *C. elegans* ciwi2 (SEQ ID NO: 6), *C. elegans* ciwi1 (SEQ ID NO: 5) and *Drosophila* aubergine (SEQ ID NO: 4). The alignment was generated using the J. Hein method with PAM250 residue weight table. Amino acid numbers for each sequence are labeled accordingly. Gaps are presented as dashes, and shaded residues match the consensus

15 exactly.

Figure 2 shows an alignment of hiwi-Hy amino acid sequence (SEQ ID NO:2) with a portion of the amino acid sequence SEQ ID NO: 9. The alignment was generated using the commercially available software BLASTP, version 2.0MP-WashU using BLAST algorithm. Differences in amino acid residues are presented as underlined amino

20 acid residues.

DETAILED DESCRIPTION OF THE INVENTION

1. DEFINITIONS

The term “primordial germ cells (PGCs)” refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal

25 ridges during embryogenesis that have the potential to differentiate in to germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived

The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes.

The term "embryonic stem cells (ES)" refers to a cell which can give rise to many differentiated cell types in an embryo or an adult, including the germ cells.

The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells which comprise the adult specialized organs, but are able to regenerate themselves.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The term "nucleotide sequence" refers to a heteropolymer of nucleotides or the sequence of these nucleotides. The terms "nucleic acid" and "polynucleotide" are also used interchangeably herein to refer to a heteropolymer of nucleotides. Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" is a stretch of polypeptide nucleotide residues which is long enough to use in polymerase chain reaction (PCR) or various hybridization procedures to identify or amplify identical or related parts of mRNA or DNA molecules.

The terms "oligonucleotides" or "nucleic acid probes" are prepared based on the polynucleotide sequences provided in the present invention. Oligonucleotides comprise portions of such a polynucleotide sequence having at least about 15 nucleotides and usually at least about 20 nucleotides. Nucleic acid probes comprise portions of such a polynucleotide sequence having fewer nucleotides than about 6 kb, usually fewer than about 1 kb. After appropriate testing to eliminate false positives, these probes may, for

example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh *et al.* [Walsh, P.S. *et al.*, *PCR Methods Appl.*, 1:241-250 (1992)].

The term "probes" includes naturally occurring or recombinant or chemically synthesized single- or double-stranded nucleic acids. They may be labeled by nick translation, Klenow fill-in reaction, PCR or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY (1992); or Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY (1989), both of which are incorporated herein by reference in their entirety.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (*e.g.*, hybridization to filter-bound DNA under in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (*e.g.*, washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances wherein hybridization of deoxyoligonucleotides is concerned, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos).

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating

sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below.

The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the term "biologically active" with reference to the stem cell maintenance factor polypeptides of the invention means that the polypeptide retains at least one of the biological activities of hiwi-Hy, preferably the ability to help maintain stem cells in a pluripotent/totipotent state, while the term "immunologically active" with reference to the stem cell maintenance factor polypeptides of the invention means that the polypeptide retains at least one of the immunologic or antigenic activities of hiwi-Hy.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, for example, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, such as stem cell maintenance factor activity, may be found by comparing the sequence of the particular polypeptide with that of homologous human or other mammalian stem cell maintenance factors and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, i.e., conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are typically in the range of about 1 to 5 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells

chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences.

Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 20% (i.e., the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.2 or less). Such a sequence is said to have 80% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, e.g., mutant, sequence of the invention varies from a listed sequence by no more than 10% (90% sequence identity); in a variation of this embodiment, by no more than 5% (95% sequence identity); and in a further variation of this embodiment, by no more than 2% (98% sequence identity). Substantially equivalent, e.g., mutant, amino acid sequences according to the invention generally have at least 95% sequence identity with a listed amino acid sequence, whereas substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (e.g., via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method.

Nucleic acid sequences encoding such substantially equivalent sequences, e.g., sequences of the recited percent identities, can routinely be isolated and identified via standard hybridization procedures well known to those of skill in the art.

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such

a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, often at least about 7 amino acids, typically at least about 9 to 13 amino acids, and, in various embodiments, at least about 17 or more amino acids. To be active, any polypeptide must have sufficient length to display biologic and/or immunologic activity.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code.

Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

The term "activated" cells as used in this application are those which are engaged in extracellular or intracellular membrane trafficking, including the export of neurosecretory or enzymatic molecules as part of a normal or disease process.

The term "purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, e.g., polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99.8% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other

component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

5 The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration.

The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed.

10 The term "intermediate fragment" means a nucleic acid between 5 and 1000 bases in length, and preferably between 10 and 40 bp in length.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation
 15 proteins secreted wholly (e.g., soluble proteins) or partially (e.g., receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R., *Cytokine*, 4(2): 134 -143
 20 (1992)] and factors released from damaged cells [e.g., Interleukin-1 Receptor Antagonist, see Arend, W.P. *et. al.*, *Annu. Rev. Immunol.*, 16:27-55 (1998)].

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

NUCLEIC ACIDS AND POLYPEPTIDES OF THE INVENTION

25 Nucleotide and amino acid sequences of the invention are reported below. Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, *et al.*, *Bio/Technology* 10, 773-778 (1992) and in R. S.

McDowell, *et al.*, *J. Amer. Chem. Soc.* 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. For example, fragments of the protein may be fused through "linker" sequences to the Fc portion of an immunoglobulin. For a bivalent form of the protein, such a fusion could be to the Fc portion of an IgG molecule. Other immunoglobulin isotypes may also be used to generate such fusions. For example, a protein-IgM fusion would generate a decavalent form of the protein of the invention.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The full-length form of the such proteins is identified in the sequence listing by translation of the nucleotide sequence of each disclosed clone. The mature form of such protein may be obtained by expression of the disclosed full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where protein of the present invention is membrane bound, soluble forms of the protein are also provided. In such forms part or all of the regions causing the protein to be membrane bound are deleted so that the protein is fully secreted from the cell in which it is expressed.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Species homologs of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species. The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides. The compositions of the present invention include isolated

polynucleotides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, novel isolated polypeptides, and antibodies that specifically recognize one or more epitopes present on such polypeptides. Species homologs (or orthologs) of the disclosed

5 polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species. The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide

10 which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

2. NUCLEIC ACIDS OF THE INVENTION

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID

15 NO: 2 or corresponding full length or mature protein or a portion thereof. A preferred nucleic acid sequence comprises the nucleic acid sequence set forth in SEQ ID NO: 1.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising the nucleotide sequence of SEQ ID NO: 1; a polynucleotide comprising the full length protein coding sequence corresponding to SEQ ID NO: 1; and a

20 polynucleotide comprising the nucleotide sequence of the mature protein coding sequence corresponding to SEQ ID NO: 1. The polynucleotides of the present invention also include, but are not limited to, polynucleotides that encode polypeptides with stem cell maintenance factor activity and that hybridize under stringent hybridization conditions to the complement of either (a) the nucleotide sequence of SEQ ID NO: 1 or (b) a nucleotide

25 sequence encoding the amino acid sequence of SEQ ID NO: 2; a polynucleotide which is an allelic variant of any polynucleotide recited above; a polynucleotide which encodes a species homolog of any of the proteins recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the stem cell maintenance factor polypeptide corresponding to SEQ ID NO: 2.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have at least about 65%, more typically at least about 70%, at least about 75%, at least about 80%, at least about 85% or at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above. The invention also provides the complement of the polynucleotides including a nucleotide sequence that has at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide encoding a polypeptide recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions which can routinely isolate polynucleotides of the desired sequence identities.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques [*see* Sambrook, J. *et al. Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY (1989)]. Useful nucleotide sequences for joining to polypeptides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic variations thereof. Allelic variations can be routinely determined by comparing the sequence provided in SEQ ID

NO: 1, or a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to SEQ ID NO: 1 with a sequence from another isolate of the same species.

To accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated. Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands).

The present invention further provides recombinant constructs comprising a nucleic acid having the sequence of SEQ ID NO: 1 or a nucleic acid encoding the full length or mature protein corresponding thereto, or a portion thereof, or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid corresponding to SEQ ID NO: 1 is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs and UMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF or UMF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman *et al.*, *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably

linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

5 Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequences that hybridize under stringent conditions to a fragment of the DNA sequence of SEQ ID NO: 1, which fragment is greater than about 10 bp, preferably 20-50 bp, greater than 100 bp, greater than 300 bp, or greater than 500 bp. In accordance with the invention, polynucleotide sequences which encode the novel nucleic acids, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. The amino acid sequence variants of the nucleic acids are preferably constructed by mutating the polynucleotide to give an amino acid sequence that does not occur in nature. These amino acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, e.g., by substituting first with conservative choices (e.g., hydrophobic amino acid to a different

hydrophobic amino acid) and then with more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells, and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein..

In a preferred method, polynucleotides encoding the novel nucleic acids are changed via site-directed mutagenesis. This method uses oligonucleotide sequences that encode the polynucleotide sequence of the desired amino acid variant, as well as a sufficient adjacent nucleotide on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman *et al.*, *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells *et al.*, *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook *et al.*, *supra*, and Current Protocols in Molecular Biology, Ausubel *et al.* Due to the inherent

degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

3. HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of stem cell maintenance factor DNA sequences allows for modification of cells to permit, or increase, expression of endogenous stem cell maintenance factors. Cells can be modified (*e.g.*, by homologous recombination) to provide increased stem cell maintenance factor expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express stem cell maintenance factor protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to stem cell maintenance factor encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (*e.g.*, *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the stem cell maintenance factor coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the stem cell maintenance factor coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation [Davis, L. *et al.*, *Basic Methods in Molecular Biology* (1986)]. The host cells containing one of polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are

usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

A number of types of cells may act as suitable host cells for expression of the protein. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from in vitro culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast, insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory

elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting, including polyadenylation signals. mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

Exemplary gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin *et al.*; International Application No. PCT/US92/09627 (WO93/09222) by Selden *et al.*; and International

Application No. PCT/US90/06436 (WO91/06667) by Skoultchi *et al.*, each of which is incorporated by reference herein in its entirety.

4. POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising the amino acid sequence of SEQ ID NO: 2 or the full length or mature protein (or a portion thereof) corresponding to the amino acid sequence encoded by the DNA of SEQ ID NO: 1. Polypeptides of the invention also include polypeptides with stem cell maintenance factor activity that are encoded by polynucleotides that (a) comprise SEQ ID NO: 1, or (b) comprise a polynucleotide encoding SEQ ID NO: 2 or (b) hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. Biologically active or immunologically active variants of the stem cell maintenance factor protein sequence corresponding to SEQ ID NO: 2 and "substantial equivalents" thereof (e.g., with 65%, 70%, 75%, 80%, 85%, 90%, typically 95%, more typically 98% or most typically 99% amino acid identity) that retain biological activity, preferably stem cell maintenance factor activity, are also contemplated. Polypeptides encoded by allelic variants may have a similar or increased or decreased activity compared to the stem cell maintenance factor corresponding to SEQ ID NO: 2.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also relates to methods for producing a polypeptide comprising growing a culture of the cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the cells or the culture medium, and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic

acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins. A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. In an alternative method, the polypeptide or protein is purified from host cells which produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. *See, e.g.,* Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, *et al.*, in *Molecular Cloning: A Laboratory Manual*; Ausubel *et al.*, *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments encoding greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention. The purified polypeptides can be used *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are collections of microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and variants (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi *et al.*, *Mol. Biotechnol.* 9(3):205-23 (1998); Hruby *et al.*, *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner *et al.*, *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in vivo tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell

cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the binding molecules may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the polypeptide of the invention or binding molecules may be complexed with imaging agents for targeting and imaging purposes.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The protein may also be produced by known conventional chemical synthesis. Methods for constructing the proteins of the present invention by synthetic means are known to those skilled in the art. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications in the peptide or DNA sequences can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (*see, e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and may thus be useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are believed to be encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, e.g., Invitrogen, San Diego, Calif., U.S.A. (the MaxBac.RTM. kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (i.e., from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl.RTM. or Cibacrom blue 3GA Sepharose.RTM.; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX). Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and In Vitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("Flag") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include hiwi-Hy analogs or variants. This embraces fragments of hiwi-Hy as well as analogs (variants) thereof in which one or more amino acids has been deleted, inserted, or substituted. Analogs of the invention also embrace fusions or modifications of hiwi-Hy wherein the protein or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to hiwi-Hy or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to desired cell types. Other moieties which may be fused to hiwi-Hy or an analog include therapeutic agents which are used for treatment of indications as described herein.

5. GENE THERAPY

Mutations in the hiwi-Hy gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal hiwi-Hy activity or to treat disease states involving hiwi-Hy (for example, to treat fertility or stem cell disorders, or to provide contraception). Delivery of a functional hiwi-Hy gene to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, *Nature*, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, *Science*, 244: 1275-1281 (1989); Verma, *Scientific American*: 68-84 (1990); and Miller, *Nature*, 357: 455-460 (1992). Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of hiwi-Hy will be useful in treating

the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of hiwi-Hy.

5.1 TRANSGENIC ANIMALS

In methods to determine biological functions of hiwi-Hy in vivo, one or more stem cell maintenance factor genes are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, *Science*, 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the role(s) hiwi-Hy play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate stem cell maintenance factor activity. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of an hiwi-Hy promoter is either activated or inactivated to alter the level of expression of the hiwi-Hy protein. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

6. USES AND BIOLOGICAL ACTIVITY

The biological activity of a polypeptide of the invention may manifest as, e.g., stem cell maintenance factor activity. The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities

described for proteins of the present invention may be provided by administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The end result desired or the mechanism underlying the particular condition or pathology will dictate whether hiwi-Hy polypeptides, binding partners thereof, or modulators (enhancers and inhibitors thereof) would be beneficial to the subject in need of treatment.

The stem cell maintenance factor hiwi-Hy is believed to play a role in the renewal and/or maintenance of stem cells, including GSCs, PGCs, ES cells and hematopoietic stem cells. Thus, administration of stem cell maintenance factor agonists is expected to promote renewal and/or maintenance of stem cells in a pluripotential or totipotential state, while administration of stem cell maintenance factor antagonists (including molecules that inhibit the expression of or activity of stem cell maintenance factors) is expected to inhibit these properties and may lead to terminal differentiation of the cells.

The hiwi gene and its homologs may also function in maintenance of asymmetric cell divisions of pluripotential stem cells in other tissues where such asymmetric cell divisions are essential to maintain progenitor populations for tissue renewal. Additionally, the ability to maintain and expand stem cell populations *in vivo* or *ex vivo* in their undifferentiated, totipotential/pluripotential state would enable the development of stable stem cell lines that could then be used to investigate early mammalian development. Such cell lines would also be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals, and in the development of biological-based sensors. Importantly, the ability to produce large quantities of human cells has important working applications for the production of substances, such as insulin or factor VIII which currently must be obtained from non-human sources or donors; implantation of cells to treat diseases such as Parkinson's; tissues for grafting such as bone marrow, skin, cartilage and bone; and organs for transplantation such as kidney and liver.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-

alpha), G-CSF, GM-CSF, thrombopoietin(TPO), (platelet factor 4 (PF-4), and basic fibroblast growth factor (bFGF).

Since stem cells can give rise to virtually any mature cell type, they are of great value for uses such as creating genetically manipulated animals, tissue and organ engineering, and cell replacement therapy. Techniques for culturing stem cells are known in the art, and once the cells are established in culture, they can be maintained using polypeptides of the invention, alone or in combination with other growth factors and/or cytokines. The resulting stem cells can be used in their existing state or can be induced to differentiate into a desired cell and/or can be mutated before administration to the subject (e.g., to render an existing gene nonfunctional or to insert another functional gene).

Alternatively, stem cells may be genetically modified to express a protein of the present invention, e.g., by introducing DNA encoding a polypeptide of the present invention which may optionally be operationally linked to an inducible expression control sequence, or by increasing expression of the endogenous stem cell maintenance factor genes. Such a construct may allow for the controlled growth and differentiation of the stem cells.

Stem cells are also useful for generating cells with multiple genetic alterations. Such genetically altered cells are desirable for many reasons, such as providing modified cells for gene therapy and replacement tissues for grafting or implantation (e.g., to avoid host rejection of the cells). This application can be used to model or treat contiguous gene disorders, aneuploidy or other large-scale chromosomal phenomenon.

The expression of the polypeptide of the present invention and its effect on the stem cells can also be manipulated to achieve a controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell population involves the use of a cell-type specific promoter driving a selectable marker (e.g., one providing resistance to an otherwise toxic drug). Under the appropriate differentiation conditions in the presence of the drug, only those cells that can activate the selectable marker (e.g., those undergoing differentiation) survive.

Enhancement of growth of stem cells such as hematopoietic stem cells, neural crest derived melanocytes, commissural axons originating from the dorsal spinal cord, crypt

cells of the gut, mesonephric and metanephric kidney tubules, and olfactory bulb is also of benefit in states where damage has occurred to these specific tissues. Use of stem cell maintenance factors is also implicated in treatment methods to reduce epithelial and endothelial proliferation or differentiation, to improve skin texture, to reduce scarring, to improve wound healing, and in other conditions associated with tissue growth as described below.

The invention also provides a cell culture comprising pluripotential stem cells and polypeptides of the invention, optionally including other exogenous growth factors and cytokines, and also provides a cell culture comprising pluripotential stem cells and other cells (e.g., fibroblasts or stromal cells) genetically altered to express polypeptides of the invention.

Inactivation of stem cell maintenance factors can also be used for contraception. Previous studies have shown that mutational inactivation of piwi gene, a member of the stem-cell maintenance factor gene family, results in the differentiation of GSCs without self-renewing divisions [Cox *et al*, *supra*]. Insertional targeted inactivation or the abrogation of the expression of the hiwi-Hy gene or its homolog or the inhibition of the functional activity of the protein encoded by the hiwi-Hy gene or its homolog can be used to inhibit GSC cell division and consequently reduce the number of mature gametes. Insertional targeted inactivation of hiwi-Hy gene or its homolog may be achieved by homologous recombination or gene therapy techniques described herein or otherwise known in the art. Alternatively, expression of hiwi-Hy gene can be controlled through triple helix formation or antisense DNA or RNA as described herein or using other methods known in the art.

6.1. RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify

chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polynucleotides of the present invention can also be expressed in stem cells by any one of the techniques commonly known to those in the art to generate stable, undifferentiated, totipotent/pluripotent stem cell lines. These stable cell lines can then serve as a source of undifferentiated totipotent / pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction based experimentation. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

The proteins provided by the present invention can similarly be used in assay to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction.

Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

5 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "*Molecular Cloning: A Laboratory Manual*", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "*Methods in Enzymology: Guide to Molecular Cloning Techniques*", Academic Press, Berger, S. L. and A. R. Kimmel eds.,
10 1987.

6.2. NUTRITIONAL USES

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a
15 source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

20 6.3. CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations.

25 6.4. IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. For example, polypeptides of the invention may be used to modulate the

immune response in the treatment of leukopaenia, immune coagulation, inflammatory reactions and autoimmune disease.

6.5. NEUROEPITHELIAL CELLS

The present invention also provides for the use of the polypeptide of the present invention in manipulating stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disorders, accidental damage, or genetic disorders. Two overlapping strategies can be used to obtain expanded populations of neuroepithelial precursor cells from stem cells using the polypeptide of the present invention: (1) the use of culture conditions effective to regulate the effect of the polypeptide of the present invention and to induce directed differentiation of the stem cells into neuroepithelial precursor cells, and (2) regulation of the effect of the polypeptide of the present invention coupled with genetic approaches to increasing the yield of neuroepithelial precursor cells. Mouse stem cells can be induced to differentiate *in vitro* with retinoic acid to form neuronal and glial precursors, positive for astrocyte (GFAP) or oligodendrocyte (O4) markers, then later into functional neurons [Fraichard *et al.*, *J. Cell Science*, 108:3161-3188 (1995)]. Cells transplanted into adult brains have been observed to innervate the host striatum [Deacon *et al.*, *Exp. Neurology*, 149:28-41 (1998)]. The present invention provides for the use of the polypeptides of the present invention for the modification and/or differentiation of stem cells for the production of neuronal stem cells in conjunction with other cytokines and/or growth factors.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head

trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

6.6.. HEMATOPOIESIS REGULATING ACTIVITY

- 5 Stem cell cultures can give rise to hematopoietic progenitor cells [Rich, *Blood*, 86:463-472 (1995)]. Stem cell derived hematopoietic cells can be generated and used to augment or replace cells damaged by illness, genetic disorder, or as alternative to the use of bone marrow transplantation when indicated. The *in vitro* derivation of hematopoietic cells from stem cells is enhanced by the addition of factors that play a major role in the self-renewal and/or maintenance of stem cells , and include but are not limited to stem cell factor (SCF), IL-3, IL-6, IL-11, GM-CSF, EPO, —CSF, G-CSF, LIF [Keller *et al.*, *Mol. Cell Biol.*, 13:473-486 (1993); Kennedy *et al.*, *Nature*, 386:488-493 (1997)]. A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid, lymphoid or erythroid cell deficiencies.
- 10
- 15 The protein of the invention may be used alone or in combination with one or more of the growth factors and/or cytokines listed above to boost or stimulate levels of hematopoietic progenitors cells in syngeneic, allogeneic, or autologous bone marrow, peripheral blood, or cord blood derived hematopoietic progenitor cell transplantation. For bone marrow or peripheral blood hematopoietic progenitor cell expansion for use in transplantation, the following three scenarios may be used alone or in combination: a donor is treated with the polypeptide of the present invention alone or in combination or with other growth factors and/or cytokines that have stem cell proliferative activity such as SCF or LIF prior to bone marrow aspiration or peripheral blood leucaphoresis to increase the number of cells available for transplantation; the bone marrow is treated *in vitro* with the polypeptide of the invention alone or in conjunction with other growth factors and/or cytokines to activate or expand the cell number prior to transplantation; and/or the recipient is treated to enhance engraftments of the donor marrow. For cord blood hematopoietic progenitor cell expansion, cord blood is contacted with the polypeptide of the present invention alone or in combination with other hematopoietic growth factors and/or cytokines. Following expansion of the cord blood hematopoietic progenitor cells,
- 20
- 25
- 30

the expanded cells are used for engraftment. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in vivo* or *ex vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The polypeptide of the present invention, alone or in combination with other growth factors and/or cytokines, may be useful for enhancing the efficiency of gene therapy based on the introduction of exogenous gene(s) into hematopoietic stem cells. Once the genes are introduced into the hematopoietic progenitor cells, the cells can then be infused in a bone marrow transplant into patients suffering from genetic disorders [Lim, *Proc. Natl. Acad. Sci. USA*, 86:8892-8896 (1989)]. Examples of genes which are useful in treating genetic disorders include genes that encode adenosine deaminase, glucocerebrosidase, hemoglobin, and cystic fibrosis chloride transporter (CFTR).

The activity of a protein of the invention on stem cells may, among other means, be measured by *in vitro* biological assays. One such biological assay is the High

Proliferative Potential Colony Forming (HPP-CFC) Cell Assay [Zsebo *et al.*, *Blood*, 73:919 (1988)]. To investigate the effect of the polypeptide of the present invention on hematopoietic progenitor cells, bone marrow cells from animals 2 days after treatment with 5-fluorouracil (5-FU) are cultured with the polypeptide of the present invention alone or in combination with other growth factors and/or cytokines in a semi-solid agar medium, and growth measured by the formation of HPP-CFC colonies in the agar [Broxmeyer *et al. Exp. Hematol.*, 5:87 (1997)].

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson *et al. Cellular Biology* 15:141-151, 1995; Keller *et al., Molecular and Cellular Biology* 13:473-486, 1993; McClanahan *et al., Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. in: *Culture of Hematopoietic Cells*. R. I. Freshney, *et al.* eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama *et al., Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In: *Culture of Hematopoietic Cells*. R. I. Freshney, *et al.* eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al., Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In: *Culture of Hematopoietic Cells*. R. I. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In: *Culture of Hematopoietic Cells*. R. I. Freshney, *et al.* eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In: *Culture of Hematopoietic Cells*. R. I. Freshney, *et al.* eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

6.7 CARDIOMYOCYTES

In yet another embodiment, stem cells generated and maintained in culture using the polypeptide(s) of the present invention can be induced to differentiate *in vitro* to form

cardiomyocytes. Directed differentiation of stem cells into cardiomyocytes may be achieved by culturing them in a culture medium containing growth factors or differentiation factors such as retinoic acid in the presence of antagonists/inhibitors of stem cell maintenance factors. Stem cell derived cardiomyocytes express appropriate cardiac-specific genes [Wobus *et al.*, *Differentiation*, 48:173-182 (1991); Klug *et al.*, *J. Clin. Invest.*, 98(1):216-224 (1998)]. Stem cell derived cardiomyocytes can be purified further for transplantation by the use of cardiomyocyte-specific promoter driving a selectable marker, e.g., X cardiac myosin heavy chain.

6.8 SKELETAL MUSCLE CELLS

In another embodiment, stem cells generated using the polypeptide of the present invention can be induced to differentiate into skeletal muscle cells *in vitro* in the presence of various growth factors and/or cytokines. For example, murine ES cells can be induced to differentiate into skeletal muscle by culture in the presence of retinoic acid. The applications of such conditions to the stem cell culture may allow for the derivation of skeletal muscle from various stem cells [Browder, L.W. In: *Principles of Tissue Engineering* eds. Lanza *et al.* Academic Press (1997)]. Alternatively, stable transfection of stem cells with genes for muscle regulatory factors (MRFs) such as Myo-D1, Myf-5 and myogenin, coupled with the inhibition of expression of endogenous stem cell maintenance factors may also allow for the generation of myocytes from stem cells [Bowder, L.W., *supra*].

6.9. TISSUE GROWTH ACTIVITY

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers, and in treatment of conditions involving hypovascularization.

A protein of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open

fracture reduction and also in the improved fixation of artificial joints. *De novo* bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

5 A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of
10 bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

 Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in
15 circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in
20 repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or
25 ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include
30 an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

Proteins of the invention may also be useful to promote better or faster closure of wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, gastric ulcers, surgical and traumatic wounds, burns and the like.

It is expected that a protein of the present invention may also exhibit activity for
 5 generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also
 10 exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting
 15 differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in:
 20 International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in:
 Winter, *Epidermal Wound Healing*, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.),
 25 Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. *Invest. Dermatol* 71:382-84 (1978).

6.10. CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example,
 30 monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or

endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes.

6.11. HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A protein of the invention may also exhibit hemostatic or thrombolytic activity. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

6.12. RECEPTOR/LIGAND ACTIVITY

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics.

By way of example, the polypeptides of the invention may be used as a ligand for a receptor thereby modulating (*i.e.*, enhancing or inhibiting) the biological activity of that receptor. Whether the polypeptides of the invention exhibit agonist, partial agonist, antagonist, or partial antagonist activity for a particular receptor, such as a stem cell maintenance factor receptor, in a particular cell type can be determined by conventional techniques known to those skilled in the art. Examples of cells that may be contacted with the protein of the invention include, but are not limited to, mammalian cells such as endothelial cells.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the invention may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego) and used in both in vivo and in vitro to bind to Tie-2.

Examples of radioisotopes include, but are not limited to, tritium and carbon-14 .

Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin. By way of example, the proteins coupled to such

5 molecules are useful in studies involving in vivo or in vitro metabolism of stem cell maintenance factors.

6.13 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and

10 totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells in vivo or ex vivo may maintain and expand cell populations in a totipotent or pluripotent state which would be useful for re-engineering damaged or diseased

15 tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels,

20 cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell

25 maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells.

Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. Furthermore, these cells can be cultured in vitro to form other differentiated cells, such as skin tissue that can be used for transplantation. In addition, the expanded stem cell populations can also be genetically

altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci, U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., *Blood*, 77: 2316-2321 (1991).

6.14 DRUG SCREENING WITH STEM CELL MAINTENANCE FACTOR POLYPEPTIDES

This invention is particularly useful for screening compounds by using the stem cell maintenance factor polypeptides of the invention, particularly binding fragments, in any of a variety of drug screening techniques. The polypeptides employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the desired

polypeptide. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention and the agent being tested or examine the diminution in complex formation
 5 between the stem cell maintenance factor polypeptides and an appropriate cell line, which are well known in the art.

6.14.1 ASSAY FOR RECEPTOR BINDING ACTIVITY

The invention also provides methods to detect specific binding of a stem cell maintenance factor of the invention to a binding partner polypeptide, and in particular a
 10 receptor polypeptide. Receptors expected to be useful in binding assays of this type may be identified using assays well known and routinely practiced in the art.

The art provides numerous assays particularly useful for identifying previously unknown binding partners for stem cell maintenance factor polypeptides of the invention. For example, expression cloning, using mammalian or bacterial cells, can be used to
 15 identify polynucleotides encoding binding partners. As another example, affinity chromatography with an immobilized stem cell maintenance factor polypeptide can be used to isolate polypeptides that recognize and bind a polypeptide of the invention. As still another example, overlay assays can be used to identify binding partner polypeptides.

6.14.2 ASSAY FOR ANTAGONISTS AND AGONISTS OF STEM CELL MAINTENANCE FACTOR ACTIVITY

Numerous techniques are known in the art to assay for agonists and antagonists of growth factors and cytokines. Such techniques can easily be adapted by one skilled in the art to identify agonists and antagonists of stem cell maintenance factor activity (which include the hiwi-Hy polypeptides of the invention, antibodies thereto, and modulators of
 25 hiwi-Hy expression or activity, e.g., antisense polynucleotides).

6.15. ANTI-INFLAMMATORY ACTIVITY

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for

example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response.

5 **6.16. LEUKEMIAS**

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic,
 10 promyelocytic, myelomonocytic, monotypic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia [for a review of such disorders, *see* Fishman *et al.*, *Medicine*, 2d Ed., J.B. Lippincott Co., Philadelphia (1985)].

15 **6.17. NERVOUS SYSTEM DISORDERS**

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration
 20 of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated
 25 with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

5 (iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

10 (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

15 (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

20 (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

25 Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or in vivo;
- (iii) increased production of a neuron-associated molecule in culture or in vivo,
- 30 e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, etc., depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

In a specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

6.18. OTHER ACTIVITIES

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition

(including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

6.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms, for example the polymorphisms illustrated below, makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as fertility disorders or stem cell disorders associated with defects in or absence of the hiwi-Hy gene) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to fertility disorders makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction

fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed.

- Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

7. THERAPEUTIC METHODS

- The novel polypeptides (including fragments, analogs and variants and antibodies) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified below.

7.1 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

- A protein of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may also contain (in addition to protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, growth factors, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, GM-CSF, thrombopoietin, stem cell factor, and erythropoietin. Compositions may include growth factors or factors that modulate angiogenesis, such as angiopoietins Ang-1, Ang-2, Ang-4, Ang-Y, and/or the

human angiopoietin-like polypeptide, and/or vascular endothelial growth factor (VEGF). Preferred growth factors for use in pharmaceutical compositions of the invention include angiogenin, bone morphogenic protein-1, bone morphogenic protein-2, bone morphogenic protein-3, bone morphogenic protein-4, bone morphogenic protein-5, bone morphogenic protein-6, bone morphogenic protein-7, bone morphogenic protein-8, bone morphogenic protein-9, bone morphogenic protein-10, bone morphogenic protein-11, bone morphogenic protein-12, bone morphogenic protein-13, bone morphogenic protein-14, bone morphogenic protein-15, bone morphogenic protein receptor IA, bone morphogenic protein receptor IB, brain derived neurotrophic factor, ciliary neurotrophic factor, ciliary neurotrophic factor receptor α cytokine-induced neutrophil chemotactic factor 1, cytokine-induced neutrophil, chemotactic factor 2 α , cytokine-induced neutrophil chemotactic factor 2 β , β endothelial cell growth factor, endothelin 1, epidermal growth factor, epithelial-derived neutrophil attractant, fibroblast growth factor 4, fibroblast growth factor 5, fibroblast growth factor 6, fibroblast growth factor 7, fibroblast growth factor 8, fibroblast growth factor 8b, fibroblast growth factor 8c, fibroblast growth factor 9, fibroblast growth factor 10, fibroblast growth factor acidic, fibroblast growth factor basic, glial cell line-derived neurotrophic factor receptor α 1, glial cell line-derived neurotrophic factor receptor α 2, growth related protein, growth related protein α , growth related protein β , growth related protein γ , heparin binding epidermal growth factor, hepatocyte growth factor, hepatocyte growth factor receptor, insulin-like growth factor I, insulin-like growth factor receptor, insulin-like growth factor II, insulin-like growth factor binding protein, keratinocyte growth factor, leukemia inhibitory factor, leukemia inhibitory factor receptor α , nerve growth factor nerve growth factor receptor, neurotrophin-3, neurotrophin-4, placenta growth factor, placenta growth factor 2, platelet-derived endothelial cell growth factor, platelet derived growth factor, platelet derived growth factor A chain, platelet derived growth factor AA, platelet derived growth factor AB, platelet derived growth factor B chain, platelet derived growth factor BB, platelet derived growth factor receptor α , platelet derived growth factor receptor β , pre-B cell growth stimulating factor, stem cell factor, stem cell factor receptor, transforming growth factor α , transforming growth factor β , transforming growth factor β 1, transforming growth factor β 1.2, transforming growth

factor β 2, transforming growth factor β 3, transforming growth factor β 5, latent transforming growth factor β 1, transforming growth factor β binding protein I, transforming growth factor β binding protein II, transforming growth factor β binding protein III, tumor necrosis factor receptor type I, tumor necrosis factor receptor type II, urokinase-type plasminogen activator receptor, vascular endothelial growth factor, and chimeric proteins and biologically or immunologically active fragments thereof.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or compliment its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein of the invention, or to minimize side effects. Conversely, protein of the present invention may be included in formulations of the particular cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent. A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein.

Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein of the present invention is administered to a mammal having a condition to be treated. Protein of the present invention may be administered in accordance with the method of the invention either alone or in
5 combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If
10 administered sequentially, the attending physician will decide on the appropriate sequence of administering protein of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

7.2. ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal,
15 transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such
20 as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in
25 fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The
30 liposomes will be targeted to and taken up selectively by the afflicted tissue.

7.3. COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein of the present invention is administered orally, protein of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein of the present invention, and preferably from about 25 to 90% protein of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein of the present invention, and preferably from about 1 to 50% protein of the present invention.

When a therapeutically effective amount of protein of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical

composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline
5 buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills,
10 dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or
15 sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium
20 alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of
25 active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or
30 magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or

liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

5 For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be
10 determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit
15 dosage form, e.g., in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous
20 solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium
25 carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

30 The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as

cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with
5 suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible
10 organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds
15 well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied;
20 other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be
25 employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the
30 compounds for a few weeks up to over 100 days. Depending on the chemical nature and

the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention. The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the

art, as disclosed, for example, in U.S. Patent. Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μ g to about 100 mg (preferably about 0.1 μ g to about 10 mg, more preferably about 0.1 μ g to about 1 mg) of protein of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon, ligament, or other tissue regeneration, the therapeutic method includes administering the composition topically, systemically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices

for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or
 5 extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in
 10 calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein
 15 compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being
 20 cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer
 25 matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the fracture repair activity of the progenitor cells. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These
 30 agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-.alpha. and TGF-.beta.),

insulin-like growth factor (IGF), other known angiopoietins, VEGF, bone morphogenic protein (BMP), as well as other cytokines and/or growth factors described herein.

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, e.g., amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (e.g., bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

7.4. EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (i.e., the concentration of the test compound which achieves a half-maximal inhibition of the C-proteinase activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD₅₀ and ED₅₀. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl *et al.*, in "*The Pharmacological Basis of Therapeutics*", Ch. 1 p.1 (1975). Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the C-proteinase inhibiting effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from in vitro data; for example, the concentration necessary to achieve 50-90% inhibition of the C-proteinase using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for the human polypeptides of the invention will be in the range of about 0.01 to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 to 25 mg/kg of patient body weight daily, varying in adults and children.

Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

7.5. PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

8. ANTIBODIES

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies include monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR and/or antigen-binding sequences, which specifically recognize a polypeptide of the invention. Preferred antibodies of the invention are human antibodies which are produced and identified according to methods described in WO93/11236, published June 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab', F(ab')₂, and F_v, are also provided by the invention. The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind hiwi-Hy polypeptides exclusively (*i.e.*, able to distinguish an hiwi-Hy polypeptide from other stem cell maintenance factor polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through

interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), *Antibodies: A*

5 *Laboratory Manual*; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the hiwi-Hy polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, hiwi-Hy polypeptides. As with antibodies that are specific for full length stem cell maintenance factor polypeptides, antibodies of the invention that
10 recognize hiwi-Hy fragments are those which can distinguish hiwi-Hy polypeptides from the family of stem cell maintenance factor polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins. Antibodies of the invention can be produced using any method well known and routinely practiced in the art.

Non-human antibodies may be humanized by any methods known in the art. In
15 one method, the non-human CDRs are inserted into a human antibody or consensus antibody framework sequence. Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or
20 quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the
25 invention are useful for detection and/or purification of the polypeptides of the invention.

Protein of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the protein. Such antibodies may be obtained using either the entire protein or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the
30 carboxyl terminus, and are conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as in R.

P. Merrifield, J. Amer. Chem. Soc. 85, 2149-2154 (1963); J. L. Krstenansky, et al., FEBS Lett. 211, 10 (1987). Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art [Campbell, A.M., *Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol.*, 35:1-21 (1990); Kohler and Milstein, *Nature*, 256:495-497 (1975)], the trioma technique, the human B-cell hybridoma technique [Kozbor *et al.*, *Immunology Today* 4:72 (1983); Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96].

Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or β -galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody

with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Research*. 175:109-124 (1988)). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A.M.,
 5 *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antiserum is isolated from the
 10 immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The present invention further provides the above-described antibodies in delectably labeled form. Antibodies can be delectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.)
 15 fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, *see* Sternberger, L.A. *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E.A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J.W. *J. Immunol. Meth.* 13:215 (1976).

20 The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as
 25 polycarbonate, complex carbohydrates such as agarose and Sepharose[®], acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. *et al.*, *Handbook of Experimental Immunology* 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. *et al.*, *Meth. Enzym.* 34 Academic Press, N.Y. (1974)). The
 30 immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in*

109201 55462001

situ assays as well as for immuno-affinity purification of the proteins of the present invention.

8.1 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the

animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a

light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that
5 binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

8.2 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No.
10 4,946,778). In addition, methods can be adapted for the construction of Fab expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal Fab fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the
15 idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F(ab')₂ fragment produced by pepsin digestion of an antibody molecule; (ii) an Fab fragment generated by reducing the disulfide bridges of an F(ab')₂ fragment; (iii) an Fab fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) Fv fragments.

8.3 BISPECIFIC ANTIBODIES

20 Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

25 Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas)

produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., 1991 EMBO J., 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol

complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the
 5 other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each
 10 Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific
 15 antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody
 20 homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making
 25 bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (VH) connected to a light-chain variable domain (VL) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the VH and VL domains of one fragment are forced to pair with the complementary VL and VH domains of another fragment, thereby forming two antigen-binding sites. Another strategy for
 30 making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc-R), such as Fc-RI (CD64), Fc-RII (CD32) and Fc-RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

8.4 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

8.5 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus

generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

8.6 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ²¹²Bi, ¹³¹I, ¹³¹In, ⁹⁰Y, and ¹⁸⁶Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as

1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

9. COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present

invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention. By providing the nucleotide sequence of SEQ ID NO: 1 or a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to SEQ ID NO: 1 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

10. TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in

length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 15241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. *Neurochem.* 56:560 (1991); Oligodeoxynucleotides as

5 Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an

10 antisense or triple helix oligonucleotide.

11. DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally

15 conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample.

20 Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise

25 contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of nucleic acid probes of the present invention and assaying for

30 binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a

container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

12. MEDICAL IMAGING

The novel polypeptides of the invention are useful in medical imaging, e.g., imaging the site of germ cell production in testes or ovaries or other sites where the stem cell maintenance factor is expressed. *See*, e.g., Kunkel *et al.*, U.S. Patent No. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide in vivo at the target site.

13. SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by the ORF from a polynucleotide of the invention to a specific domain of the polypeptide encoded by a polypeptide of the invention. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound

complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to

generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, *Application of Synthetic Peptides: Antisense Peptides*," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry* 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. *Neurochem.* 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents. Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent, in the control of bacterial infection by modulating the activity of the protein encoded by the ORF. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

14. USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from the nucleotide sequence of the SEQ ID NO: 1. Because the corresponding gene is only expressed in a limited number of tissues, especially adult tissues, a hybridization probe derived from SEQ ID NO: 1 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described US Patent Nos 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma *et al.* *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York NY (1988).

Fluorescent in situ hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data.

Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals. The nucleotide sequence may be used to produce purified polypeptides using well known methods of recombinant DNA technology.

Among the many publications that teach methods for the expression of genes after they have been isolated is Goeddel, D. *Gene Expression Technology, Methods and Enzymology*, Vol 185, Academic Press, San Diego (1990). Polypeptides may be expressed in a variety of host cells, either prokaryotic or eukaryotic. Host cells may be from the same species from which a particular polypeptide nucleotide sequence was isolated or from a different species. Advantages of producing polypeptides by recombinant DNA technology include obtaining adequate amounts of the protein for purification and the availability of simplified purification procedures.

Each sequence so obtained was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (developed by TRW Inc., Los Angeles, CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search. Peptide and protein sequence homologies were ascertained using the INHERIT™ 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology that were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

Alternatively, BLAST, which stands for Basic Local Alignment Search Tool, is used to search for local sequence alignments [Altschul, S.F., *J Mol Evol*, 36:290-300 (1993); Altschul, S.F. *et al.*, *J Mol. Biol.*, 215:403-10 (1990)]. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity.

5 Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. Whereas it is ideal for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is
10 locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database
15 sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

In addition, BLAST analysis was used to search for related molecules within the libraries of the LIFESEQ™ database. This process, an "electronic northern" analysis is
20 analogous to northern blot analysis in that it uses one cellubrevin sequence at a time to search for identical or homologous molecules at a set stringency. The stringency of the electronic northern is based on "product score". The product score is defined as (% nucleotide or amino acid [between the query and reference sequences] in Blast multiplied by the % maximum possible BLAST score [based on the lengths of query and reference
25 sequences]) divided by 100. At a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous or related molecules can be identified by selecting those which show product scores between approximately 15 and 30.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other
30 embodiments and variations may be made in the scope of the present invention.

Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples.

EXAMPLE 1

Cloning of Stem Cell Maintenance Factor cDNA

Novel nucleic acids were obtained from various cDNA libraries (prepared from human mRNA purchased from Invitrogen, San Diego, CA) using standard PCR, sequencing by hybridization (SBH) sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for pSport1 (GIBCO BRL, Grand Island, NY) vector sequences which flank the inserts.

These samples were spotted onto nylon membranes and hybridized with oligonucleotide probes to give sequence signatures. The clones were clustered into groups of similar or identical sequences, and single representative clones were selected from each group for gel sequencing. The 5' sequence of the amplified inserts was then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer.

Sequence analysis identified a polynucleotide encoding a portion of a novel polypeptide designated CG389. The 5' sequence may be determined as described in Example 2. The 3' sequence may be determined by complete sequencing of the clones listed in Table 1 or by 3' RACE using a modification of the protocol set out in example 2. The contig set forth in SEQ ID NO: 1 was deduced from clones identified below in Table 1.

Table 1

Library	No. Clones	Clone ID	Sequence ID
FLG003	1	15518182	RTA00002838F.g.13.1.P. Seq
SAL001	1	11687595	RTA00002499F.k.20.1.P.Seq

EXAMPLE 2**5' RACE Extension of hiwi-Hy Gene**

5' RACE reactions may be performed using two nested gene-specific primers (GSP) and vector primers (VP) in sequential PCR reactions on a panel of cDNA libraries. The cDNA libraries used for RACE are prepared from mRNA using a random-primed, 5' capture method to enrich for the 5' ends of genes (Carninci *et al*, *Genomics*, 37, 327-336, 1996) and cloned into the pSPORT vector (BRL Life Technologies) previously digested with NotI and SalI. The human mRNAs (Invitrogen) include message from adult brain, adult thymus, fetal muscle, fetal skin, fetal heart, fetal brain, fetal spleen, fetal liver, and fetal lung. In addition, adaptor-ligated cDNA pools (Marathon cDNAs, Clontech) made from human fetal kidney, fetal brain and adult ovary mRNAs are used in the RACE experiments.

Suitable primers may be used to amplify the desired cDNA (or genomic DNA). For example, in a first reaction, a first GSP ($T_m \sim 80^\circ\text{C}$) and VP1 ($T_m \sim 72^\circ\text{C}$) are mixed in a 5:1 ratio. Touchdown PCR is carried out as follows: an initial incubation at 96°C for one minute, followed by five cycles of 96°C for 30 seconds and 72°C for four minutes; five cycles of 96°C for 30 seconds and 70°C for four minutes; and 15 cycles of 96°C for 30 seconds and 68°C for four minutes. The products of the first reaction are diluted 1:20 and used as template for the second reaction. A second GSP and VP2 (both $T_m \sim 60^\circ\text{C}$) are mixed in a 1:1 ratio and PCR is carried out as follows: an initial incubation at 96°C for one minute; and 30 cycles of 96°C for 30 seconds, 55°C for 30 seconds, and 72°C for 90 seconds. Final RACE products are separated and identified using agarose gel electrophoresis. Selected fragments are subcloned into a TA cloning vector and the inserts are sequenced.

Table 2

pSPORT VP1: 5'AGGCACCCCAGGCTTTACACTTTA 3'	SEQ ID NO: 10
pSPORT VP2: 5'TTCCCGGGTCGACGATTTCGT 3'	SEQ ID NO: 11
Marathon cDNA VP1: 5'CCATCCTAATACGACTCACTATAGGGC3'	SEQ ID NO: 12

Marathon cDNA VP2: 5'ACTCACTATAGGGCTCGAGCGGC3'	SEQ ID NO: 13
---	---------------

EXAMPLE 3

Tissue Expression Study

5 PCR Analysis

Gene expression of the stem cell maintenance factor is analyzed using a semi-quantitative PCR-based technique. A panel of cDNA libraries derived from human tissue (from Clontech and Invitrogen) is screened with hiwi-Hy specific primers to examine the mRNA expression of hiwi-Hy in human tissues and cell types. PCR assays (For example, 94 °C for 30 sec., 58 °C for 30 sec., 72 °C for 30 sec., for 30 cycles) are performed with 20 ng of cDNA derived from human tissues and cell lines and 10 picomoles of the hiwi-Hy gene-specific primers. The PCR product is identified through gel electrophoresis. Amplified products are separated on an agarose gel, transferred and chemically linked to a nylon filter. The filter is then hybridized with a radioactively labeled (³³Pα-dCTP) double-stranded probe generated from the full-length sequence using a Klenow polymerase, random prime method. The filters are washed (high stringency) and used to expose a phosphorimaging screen for several hours. Bands of the appropriate size indicate the presence of cDNA sequences in a specific library, and thus mRNA expression in the corresponding cell type or tissue.

20 Expression analysis can also be conducted using Northern blot techniques.

EXAMPLE 4

Chromosomal Localization Study

Chromosome mapping technologies allow investigators to link genes to specific regions of chromosomes. Chromosomal mapping is performed using the NIGMS human/rodent somatic cell hybrid mapping panel as described by Drwina, H. L. et al., Genomics, 16, 311- 314, 1993 (human/rodent somatic cell hybrid mapping panel #2 purchased from the Coriell Institute for Medical Research, Camden, New Jersey). 60 ng of DNA from each sample in the panel is used as template, and 10 picomoles of the same hiwi-Hy gene-specific oligonucleotides are used as primers in a PCR assay (for example,

94°C for 30 sec., 58°C for 30 sec., 72°C for 30 sec., for 30 cycles). PCR products were analyzed by gel electrophoresis. The genomic PCR product is detected in a human/rodent somatic cell hybrid DNA containing a specific human chromosome.

EXAMPLE 5

5 Expression of Stem Cell Maintenance Factor in *E. coli*

SEQ ID NO: 1 or the full length mature protein sequence corresponding thereto is expressed in *E. coli* by subcloning the entire coding region into a prokaryotic expression vector. The expression vector (pQE16) used is from the QIAexpression[®] prokaryotic protein expression system (QIAGEN). The features of this vector that make it useful for
10 protein expression include: an efficient promoter (phage T5) to drive transcription; expression control provided by the lac operator system, which can be induced by addition of IPTG (isopropyl-β-D-thiogalactopyranoside), and an encoded His₆ tag. The latter is a stretch of 6 histidine amino acid residues which can bind very tightly to a nickel atom. The vector can be used to express a recombinant protein with a His₆ tag fused to its
15 carboxyl terminus, allowing rapid and efficient purification using Ni-coupled affinity columns.

PCR is used to amplify the coding region which is then ligated into digested pQE16 vector. The ligation product is transformed by electroporation into electrocompetent *E. coli* cells (strain M15[pREP4] from QIAGEN), and the transformed
20 cells are plated on ampicillin-containing plates. Colonies are screened for the correct insert in the proper orientation using a PCR reaction employing a gene-specific primer and a vector-specific primer. Positives are then sequenced to ensure correct orientation and sequence. To express stem cell maintenance factor polypeptides, a colony containing a correct recombinant clone is inoculated into L-Broth containing 100 μg/ml of ampicillin,
25 25 μg/ml of kanamycin, and the culture was allowed to grow overnight at 37°C. The saturated culture is then diluted 20-fold in the same medium and allowed to grow to an optical density at 600 nm of 0.5. At this point, IPTG is added to a final concentration of 1 mM to induce protein expression. The culture is allowed to grow for 5 more hours, and then the cells are harvested by centrifugation at 3000 x g for 15 minutes.

The resultant pellet is lysed using a mild, nonionic detergent in 20 mM Tris HCl (pH 7.5) (B-PER™ Reagent from Pierce), or by sonication until the turbid cell suspension turned translucent. The lysate obtained is further purified using a nickel containing column (Ni-NTA spin column from QIAGEN) under non-denaturing conditions. Briefly, the lysate is brought up to 300 mM NaCl and 10 mM imidazole and centrifuged at 700 x g through the spin column to allow the His-tagged recombinant protein to bind to the nickel column. The column is then washed twice with Wash Buffer (50 mM NaH₂PO₄, pH 8.0; 300 mM NaCl; 20 mM imidazole) and is eluted with Elution Buffer (50 mM NaH₂PO₄, pH 8.0; 300 mM NaCl; 250 mM imidazole). All the above procedures are performed at 4°C. The presence of a purified protein of the predicted size is confirmed with SDS-PAGE.

EXAMPLE 6

Evaluation of Activities *In Vitro* and *In Vivo*

The activity of stem cell maintenance factor polypeptides of the invention can be assayed using any methods known in the art, including a co-culture of stem cells with a feeder layer of stromal cells or other support cells. For example, such co-cultures could include the combination of neonatal mouse oocytes and ovarian cells. It could also comprise the co-culture of human or mouse adult bone marrow or umbilical cord blood cells, either whole or enriched stem cell populations [Zsebo *et al.*, *supra*] with stromal support cells. The stromal support cells may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts and HUVEC cells [U.S. patent No.:5, 690,926]. The ability of stem cells to propagate is assayed as described below.

This example illustrates a method for assaying the effects of stem cell maintenance factor polypeptides on stem cells. The assay measures the growth and differentiation characteristics of stem cells in the presence or absence of stem cell maintenance factor on a fibroblast feeder layer.

A supporting feeder layer is usually prepared from irradiated cells by exposing the cells to about 3,000-4,000 rads of γ -radiation [Thomson *et al.* *Proc. Natl. Acad. Sci., USA*, 92:7844-7848 (1995)]. The irradiated feeder cells are then plated in gelatinized tissue

culture plates at a density of approximately 5×10^4 cells/cm². Stem cells isolated from any one of various sources [see, e.g., Thomson *et al.*, *supra*], are cultured on the feeder layer in the presence or absence of stem cell maintenance factor polypeptide(s) and optionally one or more cytokines or growth factors listed above. The ability of stem cells to propagate *in vitro* and retain their undifferentiated characteristics is then evaluated. Cell proliferation is determined by measuring an increase in cell number and/or by measuring an increased ability of the stem cells to produce colonies in semi-solid medium *in vitro*. Stem cell numbers can also be estimated by counting the number of alkaline phosphatase positive cells [Durkova-Hills *et al.*, *Zygote*, 6:271-275 (1998)].

In vivo growth is measured by the ability of the stem cells to repopulate or reconstitute the stem cell population in an animal. Animal models known in the art such as the severe combined immunodeficiency (SCID)-mice can be used to assay the *in vivo* ability of the stem cells to repopulate and reconstitute the immune system *in vivo*.

The undifferentiated nature of the stem cells can be assayed, *in vitro* or *in vivo*, for example, by the detection of telomerase activity in stem cells. Telomerase is a marker for undifferentiation in such cells. Undifferentiated cells show high levels of telomerase activity, whereas feeder and differentiated cells typically have no detectable telomerase activity. For example, it has been shown that undifferentiated Rhesus monkey ES cells have greater than 2.5-fold level of telomerase activity in comparison with differentiated 293 cells. Such an assay may be carried out as follows.

Cell extracts of undifferentiated stem cells along with its differentiated counterpart and control telomerase-positive 293 cells are prepared by a modification of the detergent lysis method. [Kim *et al.*, *Science*, 266:2011(1994)]. Telomerase activity in the cell extracts can then be determined using a modified PCR-based TRAP assay [WO 99/20740].

The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope

of the invention are those which appear in the appended claims. All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

EXAMPLE 7

EST Sequencing of SEQ ID NO: 8 Obtained From Various Libraries

The sequence of SEQ ID NO: 8 was obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosomes using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases, RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

EXAMPLE 8

Full-length Sequence of SEQ ID NO: 8

The full length nucleic acid of SEQ ID NO: 8 was assembled from sequences that were obtained from a cDNA library by methods described in Example 7 above, and in some cases sequences obtained from one or more public databases. The nucleic acid was assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm

terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

5 Using the computer programs PHRAP (University of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against the GenBank database (i.e., dbEST version 119, gb pri 119, UniGene
10 version 119, Genepet release 119). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) or ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide and deduced amino acid sequences are shown in the Sequence Listing as SEQ ID NO: 8 and SEQ ID NO: 9.

EXAMPLE 9

15 **Expression Analysis Using SBH**

As described in Example 8, samples from over ninety cDNA libraries obtained by Hyseq were spotted onto nylon membranes and interrogated with a set of proprietary oligonucleotide probes to give clone signatures. The clones were clustered into groups of similar or identical sequences, and representative clones were selected from each group for
20 gel sequencing. Tissue expression of SEQ ID NO: 8 was determined based on the tissue source of the clones that were clustered with SEQ ID NO: 1 or with other proprietary Hyseq EST sequences used in the assemblage of SEQ ID NO: 8. Accordingly, SEQ ID NO: 8 was determined to be expressed in leukocytes, whole organ, testis, rectum, fetal lung, ovary, fetal liver-spleen, and cervix (see Table 4). The expression level of the gene,
25 represented by SEQ ID NO: 8, in each of those tissues was determined by summing the number of clones from a particular tissue in each of the clusters and dividing thus obtained sum by the total number of clones analyzed from a particular tissue. In addition, SEQ ID NO: 8 was determined to be expressed in colon tumor, ovary tumor, pancreas
adenocarcinoma, nervous tumor, fetal heart, fetal lung, testis, B-cells from tonsil,
30 fibroblasts, and fetal tissue based on public EST sequences used in the assemblage of SEQ

ID NO: 8 and their expression information as found in dbEST public database (see Table 5). The expression level of SEQ ID NO: 8 in the public domain cDNA library was determined by the total number of EST clones expressing the nucleotide sequence of SEQ ID NO:8 in the library divided by the total number of clones for the whole cDNA library.

- 5 Table 4 shows the expression data of SEQ ID NO: 8 by Screening By Hybridization obtained from the Hyseq Library containing EST sequences.

Table 4
Source Number and Tissue Distribution

Library	No. Clones	Total No. Clones	Tissue
LUC001	2	175668	Leukocyte
LUC003	2	19942	Leukocyte
SIN001	6	137888	Whole organ
ATS001	2	20952	Testis
REC001	6	27810	Rectum
FLG003	4	26606	NULL (fetal lung)
SAL001	2	36759	Whole organ
AOV001	12	256885	Ovary
FLS002	12	699888	Fetal Liver-Spleen
CVX001	2	123511	Cervix

- 20 Table 5 shows the expression data of SEQ ID NO: 8 in dbEST obtained from BLAST or other public databases.

Table 5
Source Number and Tissue Distribution

Library Identification	No. of cDNA in this library that is from this gene	Total No. Clones in this library	Library Tissue Source
NCI_CGAP_Co3	2	13335	colon tumor
NCI_CGAP_Ov8	1	454	ovary tumor

Library Identification	No. of cDNA in this library that is from this gene	Total No. Clones in this library	Library Tissue Source
NCI_CGAP_Pan1	1	24880	pancreas adenocarcinoma
NT0028	2	537	nervous tumor
Soares_fetal_h	1	38555	Soares_fetal_heart_NbHH19W
Soares_NFL_T_G	1	65657	Soares_NFL_T_GBC_S1 (fetal lung, testis, B-cells from tonsil)
Soares_senescence	1	13151	Soares_senescent_fibroblasts_NbHSF
Soares_total_f	3	27046	Soares_total_fetus_Nb2HF8_9w

EXAMPLE 11

Expression Analysis

Sequences homologous to SEQ ID NO: 8 were further obtained by a BLASTP version 2.0a1 19 MP-WashU search against Genpept release 119, using BLAST algorithm.

SEQ ID NO: 8 was determined to be expressed in adult ovary, fetal lung, fetal liver-spleen and salivary gland (Table 6). The results showed a homologue for SEQ ID NO: 8 from Genpept. The homologue, with identifiable functions for SEQ ID NO: 8, was Accession No. AF104260, *Homo sapiens* hiwi (Smithwaterman Score 1327 and 51% identity).

Table 6 shows the various tissue sources of SEQ ID NO: 8 and encoded polypeptide SEQ ID NO:9.

Table 6

Tissue Origin	RNA Source	Hyseq Library Name
Adult ovary	Invitrogen	AOV001

Fetal Lung	Invitrogen	FLG003
Fetal Liver Spleen	Colombia University	FLS002
Salivary Gland	Clonetech	SAL001

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference), SEQ ID NO: 8 was examined for domains with homology to certain peptide domains. The analysis indicated SEQ ID NO: 8 contains a Piwi domain (pFAM name), the p-value and the pFam score for the identified domain within the sequence was 5.9e-103 (p-value) and 355.4 (pFAM score).

The nucleotide sequence within the sequence that codes for signal peptide sequences and their cleavage sites can be determined using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Nielson *et al.*, In: *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites*, Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson *et al.* reference, was obtained for the polypeptide sequences.